The accelerated pace of breakthroughs in machine learning offers unparalleled optimism on the future capabilities of artificial intelligence. Despite the impressive progress, however, modern machine learning methods still operate under the fundamental assumption that the data at test time is generated by the same distribution from which training examples are collected. In order to build robust intelligent systems—self-driving vehicles, robotic assistants, smart grids—which safely interact with and control the surrounding environment, we must reason about the feedback effects of models deployed in closed-loop.

My research grapples with the challenges of using machine learning to control feedback enabled systems, grounded within the context of robotics. I use tools from control theory, statistics, and optimization to develop a principled understanding of the methods that underpin modern data-driven control pipelines. The key threads characterizing my contributions are:

- **Learning from temporally correlated data**: A distinguishing feature of feedback systems is that the data becomes correlated across time, which breaks the key independence assumption underlying machine learning methods. The existing literature suggests that learning with dependencies is harder than without, and requires assumptions that typically do not hold in practice. In [11, 12, 17], I show that for many problems the outlook is much brighter: learning from dependent data is surprisingly efficient, as if the data were actually independent.
- **Feedback induced distribution shift**: Using a model's outputs to influence future inputs induces a *distribution shift*: the distribution of inputs the model is trained on no longer reflects the test distribution. This misalignment yields errors in model predictions, which further compound at every step due to feedback. Distribution shift is a primary failure mode of imitation learning, the widely used practice of teaching robots to solve complex tasks via expert demonstrations. I show that incremental stability, a key concept from control theory, mitigates the effects of feedback induced distribution shift [10, 14]. This enables the design of practical algorithms for imitation learning that explicitly account for compounding errors in closed-loop.
- **Online adaptation to environment changes**: A deployed model should continually learn to improve online, in the face of a dynamically changing environment. In [4], I extend a classic nonlinear adaptive control method to utilize rich nonparametric function classes, removing the need for unrealistic parametric assumptions on the functional form of the environment changes. Furthermore, I provide the first rate of convergence for this classic algorithm, quantifying how fast it adapts to new conditions online [3].

As a research scientist within the robotics branch of Google Brain, my work is enriched by frequent collaborations with practitioners. To further ensure that my work is continually informed by practice, I also maintain `trajax` [7], a differentiable optimal control library for model predictive control. The `trajax` solver enables real time planning and control across a diverse set of tasks, including social navigation in indoor environments [15] and catching objects in flight [1].

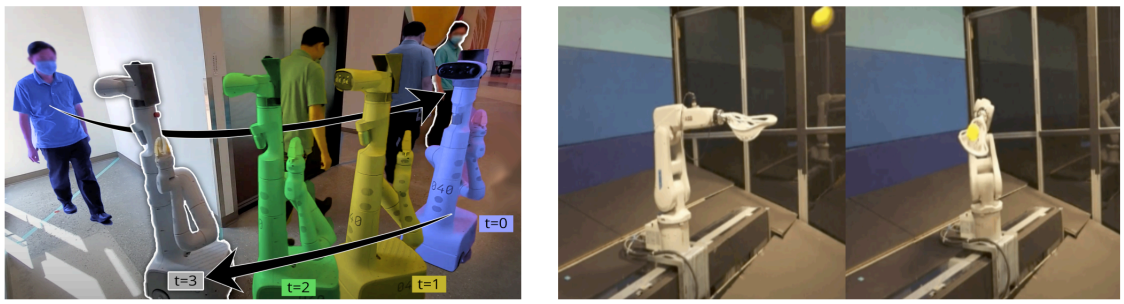

Figure 1: Robotic tasks where online planning and control are computed in real-time using `trajax`.

Looking forward, as a professor I will lead an interdisciplinary group of students to continue addressing the challenges of feedback in machine learning. I intend to focus on both the foundational aspects of using machine learning for control, and on designing practical learning algorithms to unlock new robotic capabilities while maintaining safety and robustness guarantees. Two broad future directions I intend to focus on are: (1) the foundations of learning to control from visual feedback, where the environment becomes partially observed, and (2) leveraging the recent advances in generative modeling to enable rich multi-modal robot behaviors.

Learning from temporally correlated data

A unifying theme in prior work is that the complexity of learning from a dependent process is dictated by how fast it "forgets the past", or "mixes". Specifically, data that is nearly independent across time (i.e., mixes quickly) is easier to learn from compared to data with long range temporal correlations (i.e., mixes slowly). Conversely, without some assumption limiting the dependence across time, learning is generally not possible. My work challenges this conventional wisdom.

In [12], I introduce another axis to the problem—the number of independent trajectories. This is inspired by data collection pipelines in robotics, where instead of collecting one long rollout, many independent rollouts starting from randomized initial configurations are collected. Using linear regression as a testbed, I show that by introducing this explicit "reset", the requirement that data mixes can be replaced by the more realistic requirement of having sufficiently many independent trajectories. Furthermore, the resulting risk bounds are sample efficient: for a broad class of trajectory distributions, the excess risk of linear regression from m trajectories of length T matches the optimal rate for linear regression over mT independent data points. This work yields several key takeaways for learning from dependent data: (a) mixing is not necessary for learning, and (b) even when mixing holds, it does not necessarily degrade sample complexity.

Returning to the single trajectory setting, my work [11, 17] shows that even in this more classical setup, the key takeaways from the multiple trajectories setting remain valid. In [11], I show that for linear regression problems where the covariates are generated by a linear dynamical system, a mixing assumption is unnecessary and can be replaced with a marginal stability condition. This result has seen many applications in quantifying the sample complexity of learning to control linear quadratic regulators—a classic optimal control problem—which was the focus of my dissertation [6, 9, 13]. In [17], I study a much more general class of nonparametric regression problems. I show that even when mixing holds, it only manifests through a burn-in time, after which the risk bounds for learning from a single trajectory match the optimal rates of learning from independent data.

Feedback induced distribution shift

In the context of imitation learning, one intuitively expects that expert behaviors that are more robust to small perturbations are less susceptible to error amplification through feedback. My work identifies system-theoretic properties that make this intuition precise, and proposes new algorithms that use stability theory to mitigate distribution shift.

In [14], I blend together ideas from interactive imitation learning (e.g., DAgger) and constrained policy optimization (e.g., Trust Region Policy Optimization). I show that if one combines behavior cloning with incremental stability constraints on the learned policy, then errors in policy predictions can no longer catastrophically compound due to stability. This yields the first sample complexity bounds for imitation learning that are sublinear in the task horizon length, and can become independent of the task horizon in certain cases. My work provides fine-grained insight into how the stability properties of the task reflect the hardness of imitation learning. I also use our nonlinear

Figure 2: The imitation learning pipeline. My research studies all three components.

model predictive controller (MPC) for Laikago, a quadruped robot, to study how these insights translate into practice. By taking desired walking speed as a natural proxy for task complexity, I show that to achieve a fixed level of closed-loop performance using imitation learning, more example demonstrations from the MPC controller are indeed required as the task complexity increases. This finding reflects the qualitative behavior prescribed by our theory.

In [10], I show how to construct a surrogate loss function to replace the hard incremental stability constraints enforced in [14]. Remarkably, by modifying the behavior cloning loss to include matching expert Jacobians along the demonstration trajectories, one obtains the benefits of constrained policy optimization for mitigating distribution shift, while avoiding the need for explicitly enforcing hard constraints. This yields an efficient algorithm for settings where Jacobian information is readily available, such as policy distillation, where a simple network is trained to imitate a computationally expensive controller.

Online adaptation to environment changes

A desirable property of a feedback controller is the ability to adapt online to new unforeseen changes in its surrounding environment. Adaptive control algorithms offer a promising solution, but are limited by key drawbacks when applied to modern systems.

Most adaptive control algorithms operate by making strong assumptions on both the system dynamics and the environment disturbances. Velocity gradient algorithms, a classic family of methods for nonlinear adaptive control, abstract away the details of the system dynamics via the notion of a stability (e.g., Lyapunov) function, but still make an unrealistic assumption that the disturbances modifying the environment live in the span of a known set of basis functions. In [4], I remove this assumption by showing that velocity gradient algorithms are actually compatible with modeling disturbances using rich nonparametric function classes. Furthermore, when the stability function cannot be derived from first principles, as is often the case for modern robotic systems, I show that it can instead be learned from trajectory data [2, 16].

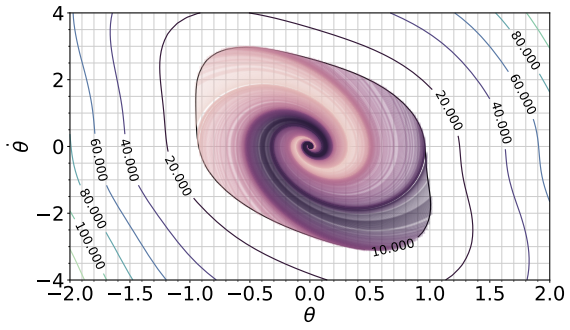

Figure 3: The level set of a Lyapunov function in phase space, learned from trajectories of a damped pendulum [2]. This function succinctly encodes the necessary information about the system dynamics for adaptive control.

Finally, while velocity gradient algorithms come equipped with convergence guarantees, these guarantees are asymptotic, and do not quantify the rate at which adaptation to new changes in the environment occurs. In [3], I provide the first finite-time analysis, using tools from online convex optimization to show that if the stability function is strongly convex, then velocity gradient algorithms actually enjoy sublinear regret.

Vision for future work

Learning to control from vision and other sensing modalities

Many robotic tasks, such as manipulating deformable objects, navigating cluttered environments, and catching balls in flight, necessitate vision in the loop. Vision, however, adds significant complexity to controller design and analysis. The two main issues are: (1) visual models must be learned from data, but quantifying the error of these models is non-trivial even without any feedback/control, and (2) most tasks requiring visual feedback are partially observed, whereas control in partially observed environments is generally intractable.

For quantifying the error of visual models, there is an exciting line of recent work on using conformal prediction methods to estimate confidence intervals. Extending these methods to apply when the inputs to the model are temporally correlated, and then using the resulting uncertainty bounds for control, is an exciting direction of future research. This aligns well with our previous work on learning control barrier functions for safe controller design from visual observations [8], assuming that valid uncertainty intervals for the perception map are provided.

On control of partially observed environments, a potential direction is to focus again on imitation learning, which allows one to sidestep hardness results. The challenge then becomes reasoning about distribution shift errors again. Formalizing incremental stability in the presence of partial observation is an exciting direction. Furthermore, there are opportunities to co-design both the data collection process and the learning algorithms. In [5], we take a first step towards this co-design: by instructing human teleoperators to purposefully inject failures, followed by demonstrating a recovery behavior, we are able to learn more robust visual policies for a t-shirt lifting task. Our final approach, however, was discovered mostly through experimentation. Characterizing the optimal teleoperation data collection strategy for a fixed operator budget is a ripe area for future research.

Applications of generative modeling to robot control

My work has thus far focused on deterministic policies—functions that take in a state vector and output a single action vector. However, for a given task configuration, there are often many equally optimal actions (e.g., many viable pick and place locations or end effector poses). Allowing policies to be inherently multi-modal yields more human-like and robust behaviors. Furthermore, multi-modality need not be limited to low level action selection, but also can be applied at higher motion and path planning levels. For the continuous spaces typical in robotics, learning a conditional distribution given an encoding of the current environment amounts to generative modeling.

The recent breakthroughs in machine learning around generative modeling provide a rich set of ideas to bring into robotics and learning to control. How to best adapt these techniques, while maintaining correctness guarantees for sampling from the true underlying distribution, is an open question. One issue of using generative models in feedback is that there are often computational constraints on inference, since decisions must be made in real-time. In contrast, modern generative models often have relatively slow sampling procedures. Understanding the fundamental tradeoffs between inference speed and sample quality is an interesting direction. Secondly, current state of the art models are quite data intensive (trained on internet scale datasets), whereas data collection is often the bottleneck in robotics. How to build data efficient models that remain expressive, while also having fast inference speed, remains open. I have begun some preliminary work, using compact energy model parameterizations to learn distributions over object grasps with noise contrastive estimation. This is only a first step, however, and I believe there is ample ground for innovation.

References

[1] S. Abeyruwan, A. Bewley, N. M. Boffi, K. Choromanski, D. D'Ambrosio, D. Jain, P. Sanketi, A. Shankar, V. Sindhwani, S. Singh, J.-J. Slotine, and S. Tu. Agile catching with whole-body mpc and blackbox policy learning. *In submission to L4DC*, 2022.

[2] N. Boffi, S. Tu, N. Matni, J.-J. Slotine, and V. Sindhwani. Learning stability certificates from data. In *Proceedings of the 2020 Conference on Robot Learning*. PMLR, 2020.

[3] N. M. Boffi, S. Tu, and J.-J. E. Slotine. Regret bounds for adaptive nonlinear control. In *Proceedings of the 3rd Conference on Learning for Dynamics and Control*. PMLR, 2021.

[4] N. M. Boffi, S. Tu, and J.-J. E. Slotine. Nonparametric adaptive control and prediction: theory and randomized algorithms. *Journal of Machine Learning Research*, 23(281):1–46, 2022.

[5] D. Brandfonbrener, S. Tu, A. Singh, S. Welker, C. Boodoo, N. Matni, and J. Varley. Visual backtracking teleoperation: A data collection protocol for offline image-based reinforcement learning. *In submission to ICRA*, 2022.

[6] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu. On the sample complexity of the linear quadratic regulator. *Foundations of Computational Mathematics*, 20(4):633–679, 2020.

[7] R. Frostig, V. Sindhwani, S. Singh, and S. Tu. trajax: differentiable optimal control on accelerators, 2021. Code available at: https://github.com/google/trajax.

[8] L. Lindemann, A. Robey, L. Jiang, S. Tu, and N. Matni. Learning robust output control barrier functions from safe expert demonstrations. *arXiv preprint arXiv:2111.09971*, 2021.

[9] H. Mania, S. Tu, and B. Recht. Certainty equivalence is efficient for linear quadratic control. In *Advances in Neural Information Processing Systems*, 2019.

[10] D. Pfrommer, T. T. Zhang, S. Tu, and N. Matni. Tasil: Taylor series imitation learning. In *Advances in Neural Information Processing Systems*, 2022.

[11] M. Simchowitz, H. Mania, S. Tu, M. I. Jordan, and B. Recht. Learning without mixing: Towards a sharp analysis of linear system identification. In *Proceedings of the 31st Conference On Learning Theory*. PMLR, 2018.

[12] S. Tu, R. Frostig, and M. Soltanolkotabi. Learning from many trajectories. *In submission to STOC*, 2022.

[13] S. Tu and B. Recht. The gap between model-based and model-free methods on the linear quadratic regulator: An asymptotic viewpoint. In *Proceedings of the Thirty-Second Conference on Learning Theory*. PMLR, 2019.

[14] S. Tu, A. Robey, T. Zhang, and N. Matni. On the sample complexity of stability constrained imitation learning. In *Proceedings of The 4th Annual Learning for Dynamics and Control Conference*. PMLR, 2022.

[15] X. Xiao, T. Zhang, K. Choromanski, E. Lee, A. Francis, J. Varley, S. Tu, S. Singh, P. Xu, F. Xia, S. M. Persson, D. Kalashnikov, L. Takayama, R. Frostig, J. Tan, C. Parada, and V. Sindhwani. Learning model predictive controllers with real-time attention for real-world navigation. In *Proceedings of the 2022 Conference on Robot Learning*. PMLR, 2022.

[16] T. Zhang, S. Tu, N. Boffi, J.-J. Slotine, and N. Matni. Adversarially robust stability certificates can be sample-efficient. In *Proceedings of The 4th Annual Learning for Dynamics and Control Conference*. PMLR, 2022.

[17] I. Ziemann and S. Tu. Learning with little mixing. In *Advances in Neural Information Processing Systems*, 2022.